Ciencia e Ingeniería
Revista Interdisciplinar de Estudios en Ciencias Básicas e Ingenierías.
Año 2016, Enero-Junio, Vol. (3) N° (1) ISSN 2389-9484.
Universidad de La Guajira, Facultades de Ciencias Básicas y Aplicadas e Ingeniería.
La Guajira-Colombia.
Revista en Línea http://revistas.uniguajira.edu.co/index.php/cei



IMPLEMENTACIÓN DE LA TÉCNICA ESTADÍSTICA BAYESIANA EN LA PLATAFORMA SOFTPROSP PARA EL CONTROL DE VARIABLES

IMPLEMENTATION OF THE BAYESIAN STATISTICAL TECHNIQUE ON THE SOFTPROSP PLATFORM FOR THE CONTROL OF VARIABLES

RAÚL J. MARTELO G., NATIVIDAD VILLABONA G., PAULO SEXTO OYOLA Q.

Programa de Ingeniería de Sistemas, Facultad de Ingeniería Universidad de Cartagena.

Recibido: Agosto 10 de 2015 Aceptado: Marzo 15 de 2016

RESUMEN

El presente proyecto tiene como objetivo implementar la técnica estadística bayesiana para el acompañamiento de estudios prospectivos en la plataforma SoftProsp. Ésta se caracteriza por ser un modelo gráfico que permite a los seres humanos tomar decisiones a partir del conocimiento y control de las variables que intervienen o influyen en el problema; se clasifican en tres (3) categorías: certeza, riesgo e incertidumbre (Beltrá, Muñoz y Muñoz, 2014).Como apoyo, se diseñaron e implementaron dos modelos, uno para especificar el procedimiento de la investigación y otro para las etapas y alcances de la técnica estadística bayesiana. En cuanto a los resultados, se pretende disminuir los inconvenientes presentados en la realización de estudios prospectivos, ampliar las opciones de utilidad de la técnica y apoyar el desarrollo de redes bayesianas con técnicas complementarias ya implementadas en la plataforma SoftProsp.

Palabras Clave: Softprosp, Estadistica Bayesisna, Variables.

ABSTRACT

The present project aims to implement Bayesian statistical technique for the accompaniment of prospective studies in the SoftProsp platform. This is characterized by being a graphic model that allows human beings to make decisions based on knowledge and control of the variables that intervene or influence the problem; (Beltrá, Muñoz and Muñoz, 2014). As support, two models were designed and implemented, one to specify the research procedure and another for the stages and scopes of the Bayesian statistical technique. As for the results, it is intended to reduce the disadvantages presented in the realization of prospective studies, to expand the utility options of the technique and to support the development of Bayesian networks with complementary techniques already implemented in the SoftProsp platform.

Key words: Softprosp, Bayesisna Statistic, Variables.

I. INTRODUCCIÓN

La técnica estadística bayesiana, también conocida como: redes causales probabilísticas, sistema expertos bayesianos, sistema expertos probabilísticos, redes causales, redes de creencia o diagramas de influencia, es una herramienta estadística aplicada en entornos bajo condiciones de incertidumbre (Ropero, Aguilera, Fernández, y Rumi, 2014) y para la toma de decisiones con base en inferencias respecto a un estado de cosas dadas (García, López, Cano, Gea, y de la Fuente, 2006).

Aunque las redes bayesianas surgen del campo de la inteligencia Artificial, gracias a su gran flexibilidad se aplican a múltiples campos y disciplinas. Díaz y Batanero (2016), complementan publicaciones de analisis de potencia o intervalos de confianza, con la metodología bayesiana y la forma en que puede complementar los analisis frecuenciales dentro de la psicología y las ciencias de la educación. Por otro lado, Delgado, Cortez e Ibáñez (2015), aplican la metodología Buchanan para la construcción de un sistema experto con redes bayesianas para apoyo al diagnóstico de la Tetralogía de Fallot en el Perú. Por ultimo, de la Torre-Gea, y otros (2014) determinan la temperatura, concentración de CO_2 y humedad relativa con respecto a la altura del cultivo, en un invernadero con ventilación natural, mediante Redes bayesianas aplicadas a un modelo de Dinámica de Fluidos Computacional.

Teniendo en cuenta lo anterior y comparando su aplicabilidad con una de las características principales de la prospectiva, la cual permite llevar a grupos humanos a tomar decisiones relevantes que los lleven a alcanzar el mejor de los futuros posibles, enfrentando situaciones de alta incertidumbre, esta técnica resulta importante en el acompañamiento de este tipo de estudios. En ella reúnen a un grupo de expertos sobre el tema, establecen un conjunto de hipótesis exhaustivas y mutuamente excluyentes, solicitan a los expertos información sobre las supuestas probabilidades basadas en su juicio o evidencia experimental y finalmente calculan las probabilidades posteriores conforme lo señalado en el Teorema de Bayes. (Miklos y Arroyo, 2008)

Es ubicada en la fase de generación, usualmente considerada el corazón del proceso, caracterizada por amalgamar, analizar, sintetizar y generar conocimiento y crear visiones e imágenes del futuro. Consta de tres (3) etapas: (a) Exploración, se comprenden los temas, problemas principales, tendencias, conductores y como los actores claves desarrollan el plan en contexto del ejercicio; (b) Análisis, se comprende como el contexto de los temas, problemas principales, tendencias y conductores influyen entre ellos, y se sintetiza el conocimiento generado en la etapa de exploración; (c) Anticipación, A partir de los análisis previos, se anticipan los futuros posibles y sugieren futuros deseables. (Popper, 2008)

No obstante, para la elaboración de estudios prospectivos se requiere de un alto presupuesto, debido a la cantidad de recursos y dificultades que se presentan (Cabarcas, Martelo y Tovar, 2013). Por estas razones, entidades como 1 LIPSOR, 2 3IE y 3 EPITA han desarrollado softwares con el propósito de mejorar la forma en que son realizados. Estas herramientas no cubren todas las falencias que se presentan en la forma tradicional de realizar prospectiva, por lo que la Universidad de Cartagena desarrollaron la plataforma SoftProsp con el propósito de cubrir algunas necesidades faltantes (Martelo, Moncaris y Velez, 2016).

SoftProsp utiliza recursos de Ingeniería de Software, Inteligencia Computacional, Sistemas de Información Geográficos y Tecnologías de la Información y las Comunicaciones, por tanto, el propósito del presente proyecto es implementar un módulo de la técnica estadística bayesiana en esta plataforma, de modo que en la realización de este tipo estudios cuenten, además de las ya implementadas, con una herramienta innovadora, capaz de presentar resultados eficaces.

II. MATERIALES Y METODOS

El presente proyecto se orienta bajo una investigación aplicada y documental. Aplicada porque se estudia un problema que conlleva al desarrollo de un aporte innovador y se están destinando esfuerzos para resolver una de tantas necesidades de la sociedad (Vargas, 2009). Documental porque en la búsqueda de respuestas específicas, se indagan teorías, libros, publicaciones científicas, entre otros documentos asociados a la técnica estadística bayesiana (Baena, 2014).

2.1 Procedimiento de la investigación

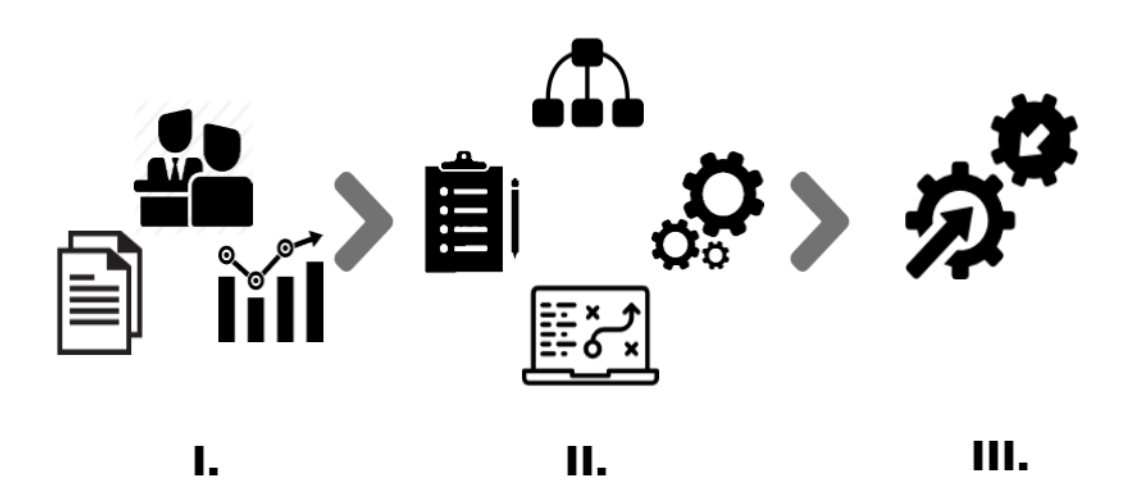


Fig. 1: Procedimiento de la investigación.

Con el propósito de implementar la técnica estadística bayesiana en la plataforma SoftProsp, se ha diseñado el modelo de la Figura 1, el cual dispone de las siguientes etapas:

2.2. Recolección de información

Para la recolección de información se han utilizado las técnicas: entrevista con el propósito de obtener información en relación con un tema determinado, buscar que la información recabada fuera los más precisa posible y aclarar los significados atribuidos por los informantes en los temas en cuestión (Díaz-Bravo, Torruco-García, Martínez-Hernández y Varela-Ruiz, 2013); y revisión y análisis documental para buscar, seleccionar, organizar y analizar un conjunto de materiales escritos que den respuesta a una o varias incógnitas (Bermeo-Yaffar, Hernández-Mosqueda yTobón-Tobón, 2016) .

2.3. Aplicación de la metodología de desarrollo RUP

Como metodología de desarrollo software se utilizará RUP (Proceso Unificado de Rational), el cual proporciona un acercamiento disciplinado a la asignación de tareas y responsabilidades en una organización de desarrollo y tiene como objetivo principal, asegurar la producción de software de alta calidad que se ajuste a las necesidades de los usuario finales con costos y calendarios predecibles. Además, se caracteriza por ser una metodología iterativa e incremental y por enfocarse en los casos de uso y la arquitectura de la solución (Martínez y Martínez, 2002).

Las fases de esta metodología de desarrollo son:

- Inicial

Se obtienen los requerimientos de la técnica estadística bayesiana. Para esto se realizan revisiones bibliográficas y entrevistas a expertos.

- Elaboración.

Abarca el diseño de modelos y diagramas UML a partir de los requerimientos establecidos.

- Construcción.

En esta fase se realiza el desarrollo de las funcionalidades, componentes y estructura de datos; también se realiza la documentación técnica a partir de los diagramas, modelos y patrones arquitectónicos establecidos en la fase anterior.

- Transición

En esta última fase se realizan las pruebas acorde a los atributos de calidad que más se ajustan al sistema y finalmente se elabora la documentación final del proyecto.

2.4 Integración de estadísticas bayesianas a SoftProsp

Culminado el proceso de desarrollo de la técnica estadística bayesiana se procede a integrar ésta con el grupo de técnicas disponibles en la plataforma SoftProsp. Para verificar el adecuado funcionamiento de estadísticas bayesianas con los otros componentes o módulos del sistema, se realizaran pruebas de integración.

III. RESULTADOS Y DISCUSIÓN

A continuación se presentan los resultados obtenidos a partir de los pasos establecidos en el procedimiento de la investigación:

3.1 Recolección de información

En esta etapa se pudo definir que estadística bayesiana es un método que permite representar relaciones de dependencia e independencia probabilística entre una colección de datos (Lacave , Molina, Redondo y Ortega, 2016). Esta es representada en un grafo acíclico dirigido en el que cada nodo representa una variable y cada arco una dependenca probabilistica en el cual se especifica la probabilidad condicional de cada variable dados sus padres; la variable a la que apunta el arco es dependiente (causa-efecto) de la que está en el origen de éste.

La topología o estructura de ésta otorga información sobre las dependencias probabilisticas entre las variables pero tambien sobre las independencias condicionales de una variable (o conjunto de variables) dada otra u otras variables; dichas independencias simplifican la representación del conocimiento (menos parámetros) y el razonamiento (propagación de las probabilidades). (Oviedo, Puris, Villacís, Delgado y Moreno, 2015)

Según lo estipulado en López (2012) y Gatica (2016), se obtuvo que la técnica estadistica bayesiana se divide en las siguientes etapas:

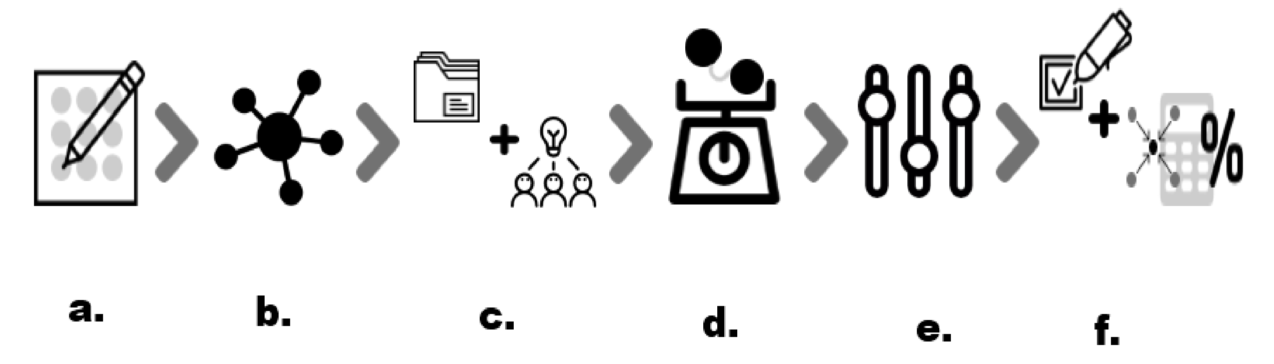


Fig. 2: Procedimiento de la técnica estadística bayesiana.

- **a.** Definición de variables: en esta etapa se deben definir las preguntas que se buscan responder, delimitando tanto los aspectos físicos como los económicos. Para luego, identificar todas las variables involucradas en el problema, sus indicadores clave, las acciones o escenarios y los datos disponibles.

- **b.** Estimación de la estructura gráfica: se deben trazar las relaciones de dependencia entre las variables "padres" e "hijos" mediante enlaces dirigidos. Hecho esto, se debe cuidar la consistencia, es decir, la lógica que presenta la red en relación al objetivo.

- **c.** Recopilación de datos: en esta etapa se incorpora la información disponible para cada una de las variables presentes en la red con el fin de ser utilizadas en el cálculo de probabilidades. Para obtener información veraz y coherente, ésta debe provenir de fuentes confiables, ejemplo de ello son: opiniones de expertos en el tema de estudio, investigaciones, informes técnicos, libros, entre otros.

- **d.** Definición de estados de las variables: para cada una de las variables se deben presentar estados, los cuales pueden corresponder a intervalos, categorías y binarios, extraídos de los datos recopilados en la etapa anterior.

- **e.** Estimación de parámetros: son introducidas las probabilidades para cada uno de los nodos (variables) a partir de los estados definidos en la etapa anterior.

- **f.** Validación y evaluación de la red: finalmente es procesada la información que contiene la red y evaluado los resultados obtenidos de acuerdo a los actores

involucrados en el problema de estudio, mediante técnicas son validados los elementos de la red.

3.2 Aplicación de la metodología de desarrollo RUP

Teniendo en cuenta lo anterior, se consideró establecer para el proceso de estadísticas bayesianas las características únicas de ésta en la plataforma SoftProsp, en donde se determinaron las siguientes mejoras frente a las soluciones existentes:

- Para la identificación de variables, la herramienta MICMAC de SoftProsp contribuye en la comprensión e identificación de problemas y representa de manera exhaustiva el sistema de variables, rediciendo así su complejidad. Ésta, con la ayuda de una matriz de análisis estructural identifica las variables y analiza las relaciones directas entre ellas.

- En relación a la etapa recopilación de datos, donde el aspecto fundamental es ampliar la información de cada una de las variables incorporadas en la red bayesiana y teniendo en cuenta el beneficio que otorga SoftProsp de manejar varias técnicas en un mismo proyecto prospectivo con el fin de solventar cada una de sus fases y verificar los datos obtenidos en cada una de estas. Como apoyo a este proceso, el módulo Indicadores bibliométricos ya implementada en la plataforma, muestra publicaciones de un área especial y la forma en que estas evolucionan en el tiempo.

- De igual manera, los módulos de Encuestas y Delphi de la plataforma, pueden ser utilizadas como herramientas para consultorías exclusivas a expertos en el tema de estudio, ejemplificar las construcción automática de redes bayesianas y recolectar información sobre puntos de vista de grupos amplios. (Popper, 2008)

- Las propuestas algorítmicas existentes para estadísticas bayesianas son diseñadas únicamente para la realización de inferencias o para el uso en el aprendizaje (Rivera, 2011). La propuesta de SoftProsp cobija ambas utilidades, ésta tendrá aplicabilidad para la sistematización, diagnostico, predicción, clasificación, entre otras.

- Utilización de minería de datos para la extracción de conocimiento a partir de bases de datos, específicamente de KDD (Knoweledge Discovery in Data Bases), con el que se identifican patrones válidos, novedosos, potencialmente útiles y entendibles sobre un conjunto de técnicas de inteligencia artificial, con el fin de encontrar conocimiento útil, valido, relevante y nuevo sobre una determinada actividad. (Lacave , Molina, Redondo y Ortega, 2016).

3.3 Integración de estadísticas bayesianas a SoftProsp

En correlación a la última etapa del procedimiento de investigación, donde se procede a integrar el módulo estadística bayesiana una vez desarrollada y probada de manera individual a la plataforma SoftProsp, con el fin de garantizar que los cambios no conlleven comportamientos no planeados o aparezcan errores adicionales, se utilizará la prueba de regresión, la cual será realizada manualmente usando herramientas de captura/reproducción automatizadas, permitiendo obtener casos de prueba y resultados para una posterior reproducción y comparación.

Se pudo observar que los sistemas probabilísticos tienden a ser muy utilizados, pero en ocasiones resulta difícil o muy dependiente de los expertos la construcción de un modelado de red bayesiana (Chávez, Grau y García, 2016). Por ello, se considera conveniente tener un método que a partir de los datos históricos permita sustituir total o parcialmente al experto en la construcción de la red y evaluar las probabilidades condicionales, en este caso el método KDD de minería de datos.
La utilización de las redes bayesianas ha sido muy difundida en diferentes campos como la manipulación de sistemas expertos en la medicina, en la depuración de programas de inteligencia artificial, en la genética, en procesos de producción; todos estos enfocados a la solución de problemas y en la identificación de probabilidades que disminuyan riesgo, como el uso primordial que se le da en la rama financiera. El uso que se le da a estas en el estudio y tratamiento de datos, consiste básicamente en mejorar el esquema de prevención de riesgo operacional, tal que sirva como sistema de ayuda en la toma de decisiones en una situación de emergencia y realizar mejoras en las estrategias operacionales. (Rivera, 2011).

IV CONCLUSIONES

La técnica estadística bayesiana de la plataforma SoftProsp al ser una aplicación web, proporciona a estudios prospectivos beneficios como la reducción de costos/tiempo tanto de los expertos para el traslado hasta el lugar del estudio, como de los interesados (organizadores) del estudio, en el alquiler de espacios y papelería. Por otro lado, la interacción de diferentes expertos, no requerirá ni de su presencia física, ni de la contratación de un traductor.

La posibilidad de complementar la estadística bayesiana con otras técnicas de SoftProsp. Indicadores bibliométricos para obtener publicaciones de un área especial y la forma en que estas evolucionan en el tiempo. MICMAC para la identificación del problema, variables y relaciones directas entre estas. Encuesta y Delphi para la recopilación de datos.

La multiplicidad de aplicación de esta técnica para distintos enfoques investigativos donde se requiera estructurar modelos de decisiones muy grandes y complejas, y además, aprender de relaciones de dependencia y causalidad, permitiendo combinar conocimiento con datos.

V. REFERENCIAS BIBLIOGRÁFICAS

Baena Paz, G. (2014). *Metodología de la investigación.* México: Grupo Editorial Patria.

Beltrán pascual, M., Muñoz Martínez, a., & Muñoz Alamillos, Á. (2014). Redes bayesianas aplicadas a problemas de credit scoring. Una aplicación práctica. *Cuadernos de Economía, 37*(104), 73-86.

Bermeo-Yaffar, F., Hernández-Mosqueda, J. S., & Tobón-Tobón, S. (2016). Análisis documental de la V Heurística mediante la Cartografía conceptual. *Ra Ximhai, 12*(6), 103-121.

Cabarcas Álvarez, A., Martelo Gómez, R. J., & Tovar Garrido, L. C. (2013). Software para mejorar la aplicación de técnicas cuantitativas en estudios prospectivos. *Cuadernos de Administración (Universidad del Valle), 29*(49), 64-74.

Chávez C, M. d., Grau Á, R., & García L, M. M. (2016). Método para construir redes bayesianas. *Revista facultad de Ingeniería*(19), 76-84.

de la Torre-Gea, G., Delfín-Santisteban, O., Torres-Pacheco, I., Soto-Zarazúa, G., Guevara-González, R., & Rico-García, E. (2014). Redes Bayesianas aplicadas a un modelo CFD del entorno de un cultivo en invernadero. *Agrociencia, 48*(3), 307-319.

Delgado Montenegro, L., Cortez Vásquez, A., & Ibáñez Prentice, E. (2015). Aplicación de metodología Buchanan para la construccion de un sistema experto con redes bayesianas para apoyo ak diagnóstico de la Tetralogía de Fallot en el Perú. *Revista de la Facultad de Ingeniería Insdutrial, 18*(1), 135-148.

Díaz, C., & Batanero, C. (2016). ¿Cómo puede el Método Bayesiano contribuir a la investigación en Psicología y Educación? *Paradigma, 27*(2), 35-53.

Díaz-Bravo, L., Torruco-García, U., Martínez-Hernández, M., & Varela-Ruiz, M. (2013). La entrevista, recurso flexible y dinámico. *Invetigación en educación médica, 2*(7), 162-167. doi:http://dx.doi.org/10.1016/S2007-5057(13)72706-6

García García, J., López Puga, J., Cano Guillén, C. J., Gea Segura, A. B., & de la Fuentes Sánchez, L. (2006). Aplicación de las redes bayesianas al modelado de las actitudes emprendedoras. *In proceedings del IV Congreso de Mtodología de Encuestas*, (págs. 235-242). Pamplona.

Gatica García, L. A. (2016). *Metodología de redes basianas para estimar el costo probable de proyectos de agua potable rural.* Undergraduate thesis, Universidad Andres Bello, Santiago de Chile.

Lacave , C., Molina, A., Redondo, M., & Ortega, M. (2016). Redes bayesianas para identifica perfiles de estudiante. Aplicación al estudio del abandono de las titulaciones de Informática en la Universidad de Castilla-La Mancha. *In Actas de las XXII JENUI* (págs. 85-92). Universidad de Almería.

López Puga, J. (2012). Cómo Construir y Validar Redes Bayesianas con Netica. *Revista Electrónica de Metodología Aplicada*, *17*(1), 1-17.

Martelo, R., Moncaris, L., & Velez, L. (2016). Integración del Ábaco de Régnier, Encuestas y Lluvia de Ideas en la Definición de Variables Claves en Estudios Prospectivos. *Información tecnológica*, *27*(5), 243-250.

Martínez, A., & Martínez, R. (17 de Diciembre de 2002). *Guía a Rational Unified Process.* Obtenido de Profesora Anaylen López: https://anaylenlopez.files.wordpress.com/2011/03/trabajo-guia20rup.pdf

Miklos, T., & Arroyo, M. (2008). *Prospectiva y escenarios para el cambio social.* México: Convenio Andrés Bello, Universidad Autónoma de México, Facultad de Ciencias Políticas y Sociales.

Oviedo, B., Puris, A., Villacís, A., Delgado, D., & Moreno, A. (2015). Análisis de datos educativos utilizando Redes Bayesianas. *In Latin American and Caribbean Conference for Engineering and Technology LACCEI.* Santo Domingo.

Popper, R. (2008). Foresight methodology. En L. Georghiou, J. Cassingena, M. Keenan, I. Miles, & R. Popper, *The Handbook the technology foresigth Concepts and Practice.* Edward Elgar Publishing Limited.

Rivera Lozano, M. (2011). *El papel de las redes bayesianas en la toma de decisiones.* Universidad del Rosario. La simulación al servicio de ka academia.

Ropero, R., Aguilera, P., Fernández, A., & Rumi, R. (2014). Redes bayesianas: una herramienta probabilistica en los modelos de distribución de especies. *Revista Ecosistemas, 23*(1), 54-60.

Vargas Cordero, Z. R. (2009). La investigación aplicada: una forma de conocer las realidades con evidencia científica. *Revista Educación, 33*(1), 155-165.